

Article

Interaction-Only Traffic Scenario Initialization for Autonomous Driving Simulation Under Strict Map-Prior Removal

Kaixi Yang *  and Shiru Qu

School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

* Correspondence: kaixiyang@mail.nwpu.edu.cn

Featured Application

This work can be applied to autonomous driving simulation systems that require traffic scenario initialization under challenging map-prior conditions, such as unavailable, delayed, incomplete, or unreliable HD-map information.

Abstract

Reliable traffic scenario initialization is essential for autonomous driving simulation, safety evaluation, and scenario completion. Most learning-based initializers rely on high-definition map priors to constrain agent placement and attribute generation. However, in practical deployment, map priors may be unavailable, incomplete, outdated, or inconsistent with temporary road changes. This study investigates traffic scenario initialization under strict map-prior removal, where both external HD-map inputs and agent-internal road-related cues are masked to reduce implicit structural leakage. We propose an interaction-only initialization framework, InterInit, which retains a fixed decoding-slot space and conditions initialization on observed agent states and multi-agent interactions. Experiments on the Waymo Open Motion Dataset first provide a vertical prior-removal analysis from full-map-prior initialization to weak-map initialization and then to strict map-prior removal. The results reveal attribute-asymmetric degradation: position and heading remain relatively stable, whereas speed becomes the dominant failure dimension. Further diagnostic analysis shows that the speed degradation is mainly caused by speed-magnitude collapse, characterized by nearly zero predicted variance rather than a large mean-speed bias. Based on this diagnosis, InterInit-SpeedFix is introduced as a localized repair strategy that models speed probabilistically in log-speed space without changing the overall architecture or inference pipeline. It reduces global Speed MMD from 0.2461 to 0.1210, corresponding to a 50.9% improvement, while maintaining stable position and heading performance. Density-stratified Speed MMD and absolute speed-magnitude error statistics further verify that the repair is consistent across different scene-density regimes and improves both typical and high-error cases. These results provide an application-oriented diagnosis-and-repair workflow for traffic scenario initialization when map priors are unavailable, and clarify the practical boundary of interaction-only initialization under strict map-prior removal.



Academic Editor: Ilse Cervantes

Received: 18 June 2026

Revised: 22 July 2026

Accepted: 25 July 2026

Published: 31 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: autonomous driving simulation; traffic scenario initialization; map-prior removal; interaction-only generation; multi-agent interaction; SpeedFix; Waymo Open Motion Dataset

1. Introduction

Traffic scenario initialization is a fundamental component of autonomous driving simulation, scenario completion, and safety-oriented evaluation. Before a simulation rollout begins, the initial states of surrounding traffic agents must be generated in a realistic and stable manner, including their positions, headings, sizes, and speeds. The quality of the initialized scenario directly affects the credibility of subsequent trajectory simulation. If the initialized position, heading, or speed distribution deviates substantially from real traffic statistics, downstream simulation may produce unrealistic interactions, abnormal collisions, or unstable rollouts, thereby weakening the reliability of safety evaluation.

In recent years, learning-based traffic scenario initialization and generation methods have made considerable progress [1–5]. Most of these methods treat high-definition (HD) maps as strong structural priors. Lane centerlines, road boundaries, topology, and candidate anchors provide explicit geometric constraints for agent placement and attribute prediction. When HD maps are complete, accurate, and up to date, these priors can substantially reduce initialization uncertainty, especially for spatial placement, heading alignment, and speed distribution. However, strong map dependence also introduces deployment risks. In real autonomous driving systems, HD maps can be locally missing, outdated, delayed in updates, or inconsistent with temporary construction zones, lane closures, traffic control, and road reconfiguration [6–9]. Therefore, traffic scenario initialization under unavailable or unreliable map priors is an important problem for practical simulation systems.

A direct strategy is to remove explicit map inputs and evaluate initialization under weak-map or explicit-map-removed settings. However, simply masking the external map tensor does not necessarily produce a true map-prior-free setting. In TrafficGen-style initialization pipelines [2], road structure may enter the model through multiple pathways. In addition to explicit HD-map inputs, agent-side features may include lane-matched vectors, local road descriptors, or global position channels that implicitly encode road geometry and topology. As a result, a model may still exploit hidden road-structure cues even when external map tensors are removed. This ambiguity motivates a stricter information-removal protocol for analyzing what multi-agent interactions alone can support after map-related priors are systematically stripped.

This paper studies traffic scenario initialization under strict map-prior removal. Different from weak-map settings that only remove explicit map inputs, the proposed protocol masks both external HD-map inputs and agent-internal road-related cues. Only observed agent kinematics, multi-agent interactions, and a fixed decoding-slot space are retained. The objective is not to replace map-conditioned initialization when reliable maps are available. Instead, this work provides a controlled framework for analyzing the capability boundary and dominant failure modes of interaction-only initialization under missing or intentionally removed map priors.

The main contributions of this study are summarized as follows:

1. A strict map-prior removal protocol is formulated for traffic scenario initialization. The protocol masks both external HD-map inputs and agent-internal road-related cues to reduce implicit structural leakage.
2. An interaction-only initialization framework, InterInit, is developed to generate traffic initial states under the strict information boundary while retaining a fixed decoding-slot space.
3. Experiments on the Waymo Open Motion Dataset show attribute-asymmetric degradation under strict map-prior removal: position and heading remain relatively stable, whereas speed becomes the dominant failure dimension.
4. Diagnostic analysis identifies speed-magnitude collapse as the primary failure mode, where the predicted speed variance approaches zero.

5. InterInit-SpeedFix is proposed as a lightweight probabilistic correction in log-speed space, reducing Speed MMD by 50.9% without changing the overall inference pipeline.

2. Related Work

2.1. Traffic Scenario Generation and Map-Prior Dependence

Traffic scenario generation is an important research direction for autonomous driving simulation, scenario completion, and safety-oriented evaluation. To construct diverse and realistic simulation environments, previous studies have developed learning-based methods for generating traffic scenes and initial agent states [1–5,10]. Many of these methods are formulated under map-conditioned or road-layout-conditioned settings, where lane centerlines, road boundaries, topology, and candidate anchors provide strong geometric constraints for agent placement and attribute prediction.

Map priors are effective because they reduce uncertainty in position, heading, and speed. Lane direction can constrain vehicle orientation, lane geometry can constrain spatial placement, and road type or local traffic structure can influence speed distributions. However, this dependence also introduces vulnerability when HD maps are missing, outdated, delayed, incomplete, or inconsistent with temporary road changes. Under such conditions, map-conditioned initializers may suffer from distribution shift and degraded robustness. This motivates the study of traffic scenario initialization under weakened or removed map priors.

2.2. Interaction Modeling Under Weak Environmental Priors

Multi-agent interaction modeling has been widely studied in trajectory prediction, traffic scene understanding, and scenario generation [11–16]. By modeling dependencies among vehicles, interaction modules can provide local relational cues such as spacing, ordering, relative motion, and motion coordination. These cues are particularly useful when environmental information is incomplete or unreliable.

However, interaction modeling does not automatically imply map-prior removal. In many trajectory prediction and scenario generation pipelines, historical motion, lane-matched vectors, local road descriptors, or global position channels may still implicitly encode road geometry and topology. Therefore, a nominal no-map setting obtained by removing only the external map tensor may still retain hidden road-related leakage through agent-side features. For strict map-prior removal, it is necessary to distinguish interaction information derived from observed agents from implicit road-related cues embedded in the input representation.

2.3. Applicability of Existing Generators, Interaction Coding, and Research Gap

Recent scenario generation methods, including SceneGen, UniGen, LCTGen, SceneControl, and Scenario Dreamer, are closely related to traffic scenario initialization and simulation-oriented scene generation [1,3–5,10]. However, these methods are generally designed under map-conditioned, road-layout-conditioned, language-conditioned, controllable, or joint map-agent generation settings. Their model interfaces often rely on lane graphs, road layouts, map-conditioned anchors, structured generation conditions, or task-specific input–output coupling. Directly removing HD-map or road-layout inputs from these models would require redesigning their input representation, assignment mechanism, decoder interface, and training protocol. The resulting variants would therefore be ad hoc reimplementations rather than faithful no-map baselines.

Another related line of work is interaction-coding-based trustworthiness testing for autonomous vehicles [11]. Such methods provide a useful perspective for collecting, coding, and constructing interactive test scenarios. However, their primary goal is to support

trustworthiness-oriented scenario testing through interaction coding and data-collection protocols, whereas the present study focuses on traffic scenario initialization under a strict information-removal boundary. In particular, our objective is to examine what remains learnable when both explicit HD-map inputs and implicit agent-internal road-related cues are removed. Therefore, interaction coding is discussed as related work rather than adopted as the main generation protocol in this study.

Accordingly, this study uses controlled TrafficGen-family baselines under a consistent evaluation pipeline. This design allows the comparison to distinguish full-map-prior initialization, test-time explicit-map removal from full-map-trained models, map-light train/test settings, and strict map-prior removal. Existing studies have not sufficiently separated these information regimes. When only external map tensors are removed, the model may still recover road structure from agent-internal road cues or may suffer from train-test prior mismatch. This ambiguity motivates the strict map-prior removal protocol used in this study.

3. Materials and Methods

3.1. Problem Definition

Let $A = \{a_i\}_{i=1}^N$ denote the set of observed agents in a traffic scenario, and let $C = \{c_j\}_{j=1}^M$ denote the set of fixed decoding slots. The goal of traffic scenario initialization is to generate the initial-state distribution of target traffic agents under removed or unavailable map priors. The initialized scenario y includes position, size, heading, speed, and velocity direction.

The conditional distribution studied in this work is

$$p(y \mid A, C), \quad (1)$$

where initialization is conditioned on observed agent states and fixed decoding slots, but not on explicit HD-map inputs or agent-internal road-related cues. The fixed decoding slots C are used to define the output space and supervision assignment. They are not provided to the encoder as explicit road geometry or lane topology.

This formulation separates the initialization problem from map-conditioned candidate placement. Instead of asking whether complete map priors improve initialization, this work asks what remains learnable when map-related information pathways are systematically removed and only agent observations, interaction context, and a fixed output-slot space remain available.

3.2. Information Decomposition for Map-Prior Removal

Traffic scenario initialization is usually supported by multiple types of information. In map-conditioned methods, the generated initial states are constrained not only by observed agents, but also by explicit road geometry and implicit road-related attributes. To define a strict no-map setting, these information sources must be separated before any masking operation is applied.

In this study, the available information is decomposed into five categories. The first category is the external HD-map prior, including map polylines, lane boundaries, lane centerlines, road topology, and map-conditioned candidate anchors. These inputs provide explicit geometric and topological constraints. The second category is agent-internal road-related information, such as lane-matched vectors, local road descriptors, or global position channels that may implicitly encode map structure. Although these variables may appear as agent-side features, they can still leak road information into the model. The third category is observed agent kinematics, including available agent states that describe existing traffic

participants. The fourth category is agent–agent interaction information, which captures spatial and behavioral dependencies among observed agents. The fifth category is the fixed decoding-slot space, which defines where predictions are assigned but is not treated as an explicit map representation.

This decomposition is important because explicit map removal alone does not necessarily produce a strict no-map initializer. A model may still infer road structure through agent-internal road-related cues even when external map tensors are removed. Therefore, strict map-prior removal should be understood as the removal of both explicit HD-map inputs and implicit road-related feature channels, rather than merely deleting a map tensor from the input.

3.3. Strict Map-Prior Removal Protocol

Based on the above information decomposition, the strict map-prior removal protocol defines the core information boundary of this study. The protocol removes external map polylines, lane topology, map-conditioned candidate information, and agent-internal road-related cues. It retains only observed agent kinematics, multi-agent interaction context, and fixed decoding slots.

Let the original input representation be denoted as

$$X = \{X_A, X_C, X_M, X_Q\},$$

where X_A denotes observed agent-state features, X_C denotes fixed decoding-slot features, X_M denotes external HD-map features, and X_Q denotes agent-internal road-related cues. The strict map-prior removal protocol can be represented by a projection operator:

$$\tilde{X} = \mathcal{P}_{\text{strict}}(X) = \{X_A, X_C, \mathbf{0}, \mathbf{0}\}. \quad (2)$$

This operation removes both explicit map features and implicit road-related agent-side cues, while preserving the observed agent states and fixed decoding slots required for controlled initialization.

The purpose of this protocol is not to remove all spatial indexing from the task. The fixed decoding slots are retained because the initialization problem still requires a controlled output space and a consistent supervision assignment mechanism. However, these slots are not supplied to the encoder as explicit road geometry, lane topology, or map-conditioned anchors. In this sense, the proposed setting should be interpreted as strict removal of map-prior inputs, rather than complete removal of all spatial organization.

Formally, the retained information can be written as

$$\mathcal{I}_{\text{retain}} = \{A, C, \mathcal{R}_{AA}\}, \quad (3)$$

where A denotes observed agent states, C denotes the fixed decoding-slot space, and $\mathcal{R}_{AA}$ denotes agent–agent interaction relations derived from the observed agents. The removed information can be written as

$$\mathcal{I}_{\text{remove}} = \{\mathcal{M}_{\text{HD}}, \mathcal{G}_{\text{lane}}, \mathcal{Q}_{\text{road}}\}, \quad (4)$$

where $\mathcal{M}_{\text{HD}}$ denotes external HD-map inputs, $\mathcal{G}_{\text{lane}}$ denotes lane geometry and topology, and $\mathcal{Q}_{\text{road}}$ denotes agent-internal road-related cues. The model is therefore trained and evaluated under the restricted condition

$$p(y \mid \mathcal{I}_{\text{retain}}), \quad \mathcal{I}_{\text{remove}} \notin \text{Input}. \quad (5)$$

This protocol makes the information boundary explicit and reduces the risk that performance under a nominal no-map setting is supported by hidden map-related leakage.

3.4. Vector-Based Feature Representation and Masking Details

As illustrated in Figure 1, the strict map-prior removal protocol separates the original representation into masked and retained information components. Figure 1 clarifies which blocks are removed under InterInit and summarizes the retained pairwise interaction descriptors. The detailed feature-level masking policy is summarized later in Table 1.

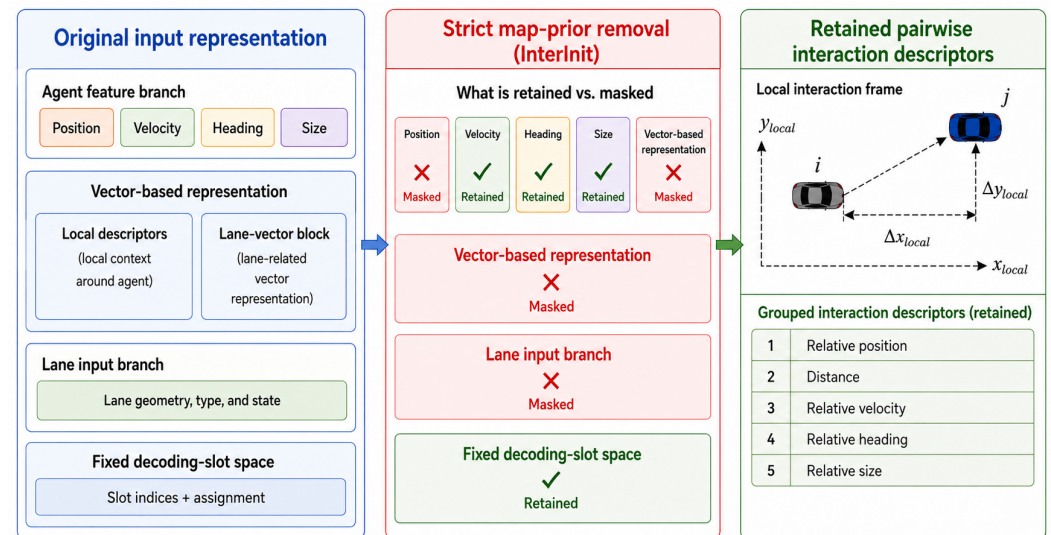


Figure 1. Simplified information-masking diagram under strict map-prior removal. The figure summarizes the original input representation, the retained and masked components under InterInit, and a grouped view of the retained pairwise interaction descriptors. Ego-relative position, agent-internal vector-based road-related representations, and the lane input branch are masked, while velocity, heading, size, pairwise interaction descriptors, and the fixed decoding-slot space are retained. The interaction descriptors are grouped for readability in this figure; their full 10-dimensional definition is given in Equation (9), and detailed feature-level masking is summarized in Table 1.

Table 1. Raw input features and masking policy under strict map-prior removal.

Information Category	Input Field	Index/Dimension	Concrete Feature Content	InterInit Treatment
External HD-map prior	lane_inp	indices 0–3	Lane polyline geometry	Masked
External HD-map prior	lane_inp	index 4	Lane type	Masked
External HD-map prior	lane_inp	index 5	Traffic state/ traffic-light state	Masked
Agent-internal road-related information	agent_feat	indices 0–1	Ego-relative position	Masked
Agent-internal vector-based representation	agent_feat	indices 8–12	long_perc, lat_perc, v_value, v_relative_dir, and relative_dir	Masked
Agent-internal road-related information	agent_feat	indices 13–18	Lane-vector block from vec_based_rep	Masked
Observed agent kinematics	agent_feat	indices 2–3	Velocity (v_x, v_y)	Retained
Observed agent kinematics	agent_feat	indices 4–5	Heading representation ($\cos \theta, \sin \theta$)	Retained
Observed agent attributes	agent_feat	indices 6–7	Length and width	Retained

The left block of Figure 1 shows the original representation used before masking. The agent feature branch contains position, velocity, heading, size, and vector-based representations, while the lane input branch contains lane geometry, lane type, and traffic-state information. The fixed decoding-slot space is shown as a separate retained component because it defines slot indices and supervision assignment, rather than serving as a lane-input feature channel. The middle block shows the strict masking result: position, vector-based representations, and the lane input branch are removed, whereas velocity, heading, size, and the fixed decoding-slot space are retained. The right block gives a grouped summary of the retained pairwise interaction descriptors. Their full 10-dimensional definition is provided in Equation (9). Within InterInit, these retained descriptors provide local agent–agent relational cues for interaction aggregation before slot-conditioned initialization, rather than serving as map-derived structural inputs.

The original feature input mainly consists of two model-input branches: the agent feature branch `agent_feat` and the lane input branch `lane_inp`. The former has 19 dimensions for each agent, while the latter has 6 dimensions for each lane element. In addition, the fixed decoding-slot space defines slot indices and supervision assignment, but it is not treated as a raw feature channel in either `agent_feat` or `lane_inp`.

The 19-dimensional `agent_feat` is formed by concatenating agent position, velocity, heading, size, and an 11-dimensional vector-based representation:

$$\text{agent_feat} = [\mathbf{p}, \mathbf{v}, \mathbf{h}, \mathbf{s}, \mathbf{r}], \quad (6)$$

where $\mathbf{p} \in \mathbb{R}^2$ denotes ego-relative position, $\mathbf{v} \in \mathbb{R}^2$ denotes velocity, $\mathbf{h} \in \mathbb{R}^2$ denotes the cosine/sine heading representation, $\mathbf{s} \in \mathbb{R}^2$ denotes length and width, and $\mathbf{r} \in \mathbb{R}^{11}$ denotes the vector-based representation. The vector-based representation is further divided into two parts:

$$\mathbf{r} = [\mathbf{r}_{\text{local}}, \mathbf{r}_{\text{lane}}], \quad (7)$$

where $\mathbf{r}_{\text{local}} \in \mathbb{R}^5$ contains `long_perc`, `lat_perc`, `v_value`, `v_relative_dir`, and `relative_dir`, while $\mathbf{r}_{\text{lane}} \in \mathbb{R}^6$ corresponds to the lane-vector block. Under the strict no-map setting, both the local vector descriptor block and the lane-vector block are masked in `agent_feat`. The 6-dimensional `lane_inp` contains explicit lane/map information, including lane polyline geometry, lane type, and traffic state, and is fully masked.

Under InterInit, the tensor shapes are kept unchanged, but all dimensions carrying external structural information or agent-internal road-related vector information are set to zero. Specifically, the raw-feature masking operation is implemented as

$$\begin{aligned} \text{agent_feat}[\dots, 0:2] &= 0, \\ \text{agent_feat}[\dots, 8:19] &= 0, \\ \text{lane_inp}[\dots, :] &= 0. \end{aligned} \quad (8)$$

The slice notation follows zero-based half-open indexing, where `[a:b]` denotes indices from `a` to `b − 1`.

The first operation masks ego-relative position, which may encode strong spatial structural priors. The second operation masks the full 11-dimensional vector-based representation embedded in `agent_feat`, including the local descriptor block and the lane-vector block. The third operation masks the entire lane input branch, including lane geometry, lane type, and traffic state. Consequently, `agent_feat` remains 19-dimensional, but only 6 effective non-zero raw agent-feature dimensions are retained: velocity, heading, and size.

Although the raw position and vector-based representation in `agent_feat` are masked, the interaction module still receives pairwise interaction descriptors computed between observed agents. For an ordered pair of agents (i, j) , the interaction descriptor is defined as

$$\mathbf{e}_{ij} = [\Delta x_{ij}^{\text{local}}, \Delta y_{ij}^{\text{local}}, d_{ij}, \Delta v_{ij}^x, \Delta v_{ij}^y, \cos(\Delta \theta_{ij}), \sin(\Delta \theta_{ij}), \Delta l_{ij}, \Delta w_{ij}]. \quad (9)$$

Here, $\Delta x_{ij}^{\text{local}}$ and $\Delta y_{ij}^{\text{local}}$ denote the pairwise relative position in the local agent-interaction frame, and $d_{ij} = \sqrt{(\Delta x_{ij}^{\text{local}})^2 + (\Delta y_{ij}^{\text{local}})^2}$ denotes the corresponding pairwise distance. The remaining terms are derived from the retained 6-dimensional agent state, including velocity, heading, and size: $(\Delta v_{ij}^x, \Delta v_{ij}^y)$ denotes relative velocity, Δv_{ij} denotes relative speed, $\cos(\Delta \theta_{ij})$ and $\sin(\Delta \theta_{ij})$ denote relative heading, and $(\Delta l_{ij}, \Delta w_{ij})$ denote relative length and width.

This design separates raw map-related feature masking from interaction modeling. The pairwise relative-position terms are used only to construct agent–agent interaction relations and do not contain HD-map polylines, lane topology, lane centerlines, lane type, traffic-light state, or map-matched lane directions. Therefore, they are treated as interaction descriptors rather than road-related priors. The strict no-map setting thus removes explicit and implicit road-structure cues while retaining the relative information required for multi-agent interaction modeling.

Consistent with the five-category information decomposition illustrated in Figure 1, Table 1 summarizes the corresponding raw input features and the masking policy used in `InterInit`.

The pairwise interaction descriptor consists of relative position, distance, velocity, speed, heading, and size terms, as defined in Equation (9).

The fixed decoding-slot space is not listed as a raw input feature in Table 1, because it defines the output slots and supervision assignment structure rather than a feature channel in `agent_feat` or `lane_inp`. It is retained without being used as lane geometry, road topology, or map-conditioned candidate information.

3.5. Interaction-Only Initialization Principle

Under strict map-prior removal, the model can no longer rely on lane geometry, road topology, or map-conditioned candidate anchors. Initialization must instead be supported by observed agent states, agent–agent relations, and the fixed decoding-slot structure. This leads to an interaction-only initialization principle: the model should infer plausible initial attributes from the distribution of observed traffic participants and their mutual spatial relations, rather than from explicit road priors.

This principle does not imply that all initialization attributes are equally recoverable without map priors. Spatial attributes may still be partially constrained by observed traffic configurations, local density patterns, relative positions, and interaction structure. For example, the arrangement of existing agents may provide weak cues about occupied regions, plausible gaps, and local orientation consistency. In contrast, dynamic attributes may depend more strongly on road semantics, lane type, curvature, speed limits, intersection structure, and local traffic rules. These factors are usually encoded by map-related priors and may become difficult to infer after strict removal.

Therefore, the proposed framework is not designed to prove that map priors are unnecessary. Instead, it provides a controlled setting for identifying which parts of traffic scenario initialization remain supported by interaction-only information and which parts expose dominant failure modes after map-prior removal. This perspective turns no-map initialization from a simple input-masking operation into an information-boundary analysis problem.

Figure 2 summarizes the overall workflow of this study, including strict map-prior removal, interaction-only initialization, and the later diagnosis-and-repair process.

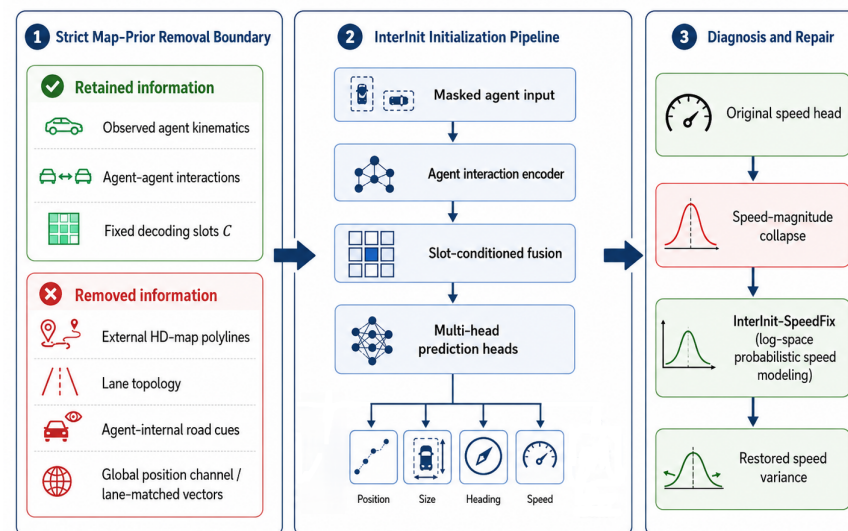


Figure 2. Overall study workflow for strict map-prior removal, InterInit initialization, and post-diagnostic SpeedFix repair. The left panel defines the retained and removed information under the strict no-map setting. The middle panel shows the original InterInit pipeline for interaction-only initialization. The right panel summarizes the later diagnosis-and-repair process: after speed-magnitude collapse is identified from the strict no-map results, InterInit-SpeedFix is introduced as a localized repair strategy for the diagnosed speed failure.

3.6. InterInit Initialization Pipeline

InterInit performs interaction-only initialization under the strict map-prior removal boundary. The model first encodes masked agent inputs and aggregates multi-agent interaction context. The interaction representation is then fused with the fixed decoding-slot representation. Finally, multiple prediction heads output slot-selection probability, position, size, heading, speed, and velocity direction.

The pipeline can be summarized as

$$h_a = f_{\text{agent}}(\tilde{x}_a), \quad (10)$$

$$h_c = f_{\text{center}}(\tilde{x}_c, h_a), \quad (11)$$

$$\hat{y} = g(h_c), \quad (12)$$

where $f_{\text{agent}}(\cdot)$ denotes the agent encoder and interaction aggregation module, $f_{\text{center}}(\cdot)$ denotes the slot-conditioned fusion module, and $g(\cdot)$ denotes the multi-head prediction module. The variables $\tilde{x}_a$ and $\tilde{x}_c$ indicate that corresponding input channels are masked according to the strict map-prior removal protocol.

The initialized output of each decoding slot is represented as a multi-attribute state:

$$\hat{y}_j = \{\hat{o}_j, \hat{p}_j, \hat{s}_j, \hat{\theta}_j, \hat{v}_j, \hat{d}_j\}, \quad j = 1, \dots, M, \quad (13)$$

where $\hat{o}_j$ denotes the slot-selection probability, while $\hat{p}_j$, $\hat{s}_j$, $\hat{\theta}_j$, $\hat{v}_j$, and $\hat{d}_j$ denote position, size, heading, speed magnitude, and velocity direction, respectively. The slot-selection probability is used to determine whether a decoding slot corresponds to a valid generated

agent, whereas the remaining heads describe the initialized agent attributes. The interaction-only initializer can then be written as

$$\hat{Y} = F_{\theta}(\tilde{X}) = \{\hat{y}_j\}_{j=1}^M, \quad (14)$$

where $\tilde{X} = \mathcal{P}_{\text{strict}}(X)$ is the strictly masked input representation. Compared with full-map-prior map-conditioned models, the key difference of InterInit is not a stronger map-conditioned backbone, but the controlled removal of map-related information pathways.

InterInit adopts fixed-slot parallel initialization rather than autoregressive actor insertion. Each decoding slot predicts a multi-attribute initial state under a consistent output-space and supervision-assignment protocol. Replacing this formulation with an LSTM-based autoregressive sampler would introduce a different model family and generation protocol, which is outside the controlled prior-removal comparison considered in this study.

3.7. Multi-Head Prediction and Training Objective

For each fused slot representation, the model predicts slot-selection probability, position distribution, size distribution, heading distribution, speed, and velocity direction. The original InterInit training objective consists of multiple attribute-specific losses:

$$\mathcal{L} = \mathcal{L}_{\text{prob}} + \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{bbox}} + \mathcal{L}_{\text{heading}} + \mathcal{L}_{\text{speed}} + \mathcal{L}_{\text{vdir}}. \quad (15)$$

The slot-selection head uses binary cross-entropy. The position, size, and heading heads use negative log-likelihood losses. The original speed head uses deterministic regression, and the velocity-direction head uses an L1 loss. This objective corresponds to the multi-attribute nature of traffic scenario initialization. The diagnostic repair for the speed head is not introduced at this stage; it is proposed later after the dominant speed-related failure mode is identified from the strict no-map results.

4. Experimental Setup

4.1. Dataset and Task

Experiments are conducted on the Waymo Open Motion Dataset [17]. The task is traffic scenario initialization: given the ego vehicle and a small set of observed context agents, the model generates initial states for the remaining traffic agents. The evaluation uses 10,000 test scenarios, corresponding to more than 60,000 samples.

4.2. Compared Methods

The compared methods are organized by map-prior information regimes rather than as direct strict no-map variants of all existing generators. As discussed in Section 2.3, constructing strict no-map versions of recent map-conditioned or joint map-agent generators would require changing their original input interfaces, assignment mechanisms, decoder designs, and training protocols. Therefore, this study adopts controlled TrafficGen-family baselines to analyze performance changes across full-map-prior, test-time explicit-map-removal, map-light, and strict no-map settings.

TrafficGen and TraffusionGen are used as full-map-prior references. TrafficGen represents full-map-prior initialization without enhanced internal interaction modeling. TraffusionGen is an interaction-augmented TrafficGen-style initializer that preserves the original structural pathway for map-conditioned scene-layout encoding, while introducing an explicit interaction pathway inside the backbone. This interaction pathway operates on the same vectorized scene representation and uses multi-head self-attention to model dependencies among scene elements, including agent spacing, ordering, and motion coordination. The structural and interaction features are fused in the encoder through an MLP-based

fusion module with normalization, nonlinear activation, and residual connection, and a shared decoder predicts the final initialization from the fused representation. Thus, TraffusionGen serves as the full-map-prior reference with enhanced interaction modeling, whereas InterInit is evaluated under strict map-prior removal. TrafficGen-WeakMap and TraffusionGen-WeakMap are used as test-time explicit-map-removal sensitivity baselines. They are trained under full-map-prior conditions, but the explicit HD-map branch is removed at test time. They are not strict no-map methods, because their training procedures and remaining agent-side channels may still be coupled with map-conditioned representations or agent-internal road-related cues. Their role is to measure train–test prior mismatch under missing explicit maps.

MapLight-AgentInit is included as a dedicated map-light baseline. It is trained and evaluated after removing explicit global HD-map inputs, while retaining bounded agent-centered local segment cues. InterInit further removes both explicit HD-map inputs and agent-internal road-related cues during both training and testing, while retaining observed agent kinematics, agent–agent interaction information, and the fixed decoding-slot space. InterInit-SpeedFix uses the same strict no-map information boundary and adds a localized probabilistic speed repair after speed-magnitude collapse is diagnosed. Table 2 summarizes the compared methods and their corresponding roles.

Table 2. Comparison of the evaluated initialization methods and their information regimes.

Method	Prior Condition	Train/Test Protocol	Interaction	Role
TrafficGen	Full-map prior	Full-map train/test	Not enhanced	Full-map-prior baseline
TraffusionGen	Full-map prior + interaction	Full-map train/test	Enhanced	Full-map-prior interaction reference
TrafficGen-WeakMap	Test-time explicit-map removal	Full-map train/map-removed test	Not enhanced	Explicit-map-removal sensitivity
TraffusionGen-WeakMap	Test-time explicit-map removal + interaction	Full-map train/map-removed test	Enhanced	Interaction-enhanced map-removal sensitivity
MapLight-AgentInit	Map-light	Explicit-map-removed train/test with local cues	Preserved	Map-light train/test baseline
InterInit	Strict no-map	Strict no-map train/test	Preserved	Strict map-prior removal method
InterInit-SpeedFix	Strict no-map + SpeedFix	Strict no-map train/test + speed repair	Preserved	Repaired strict no-map variant

4.3. Evaluation Metrics

Maximum Mean Discrepancy (MMD) is used to measure distributional discrepancy between generated and ground-truth initialization attributes. Lower MMD indicates better distributional alignment. Four initialization attributes are evaluated: Speed, Position, Heading, and Size. These attributes correspond to complementary aspects of traffic scenario initialization. Speed measures whether the initialized agents follow realistic motion-intensity distributions, which is important for subsequent simulation rollout and interaction dynamics. Position evaluates whether agents are placed in statistically plausible spatial locations. Heading reflects whether the initialized orientation is consistent with local traffic structure and agent motion direction. Size measures whether the generated vehicle dimensions match the real distribution, which affects occupancy, spacing, and potential

overlap in the initialized scene. Together, these four attributes evaluate both the spatial plausibility and dynamic consistency of the initialized traffic scenario.

In addition to global MMD, diagnostic metrics are used after a dominant failure dimension is identified. These diagnostic metrics include attribute-level mean, attribute-level standard deviation, density-stratified MMD, and absolute error statistics. They are used to determine whether the observed degradation comes from mean shift, variance collapse, density-dependent instability, or high-error samples.

For downstream trajectory-generation validation, Mean ADE, Mean FDE, and Scenario Collision Rate (SCR) are further reported. ADE measures the average Euclidean displacement error over the generated trajectory, and FDE measures the displacement error at the final prediction step. SCR measures the percentage of generated scenarios that contain collision events during downstream trajectory generation. These metrics are used only for the downstream validation of InterInit and InterInit-SpeedFix, while the main initialization evaluation remains based on MMD.

4.4. Implementation Details

All methods are evaluated under the same data split and evaluation pipeline. Training uses 100 epochs, batch size 32, learning rate 3×10^{-4} , weight decay 4×10^{-4} , and 10 data-loading workers. Experiments are conducted on a single RTX 4090 GPU. For InterInit and the post-diagnostic repaired variant reported later, multiple random seeds are used and results are reported as mean $\pm$ standard deviation.

5. Results

5.1. Global Initialization Performance Under Progressive Map-Prior Removal

Table 3 reports the global MMD results under the progressive information settings defined in Table 2. The comparison includes full-map-prior references, test-time explicit-map-removal baselines, a dedicated map-light baseline, and the proposed strict no-map initializer. Lower MMD values indicate better distributional alignment with the ground-truth initialization attributes.

Table 3. Global MMD comparison under progressive map-prior removal. Lower values are better.

Model	Prior Condition	Speed	Position	Heading	Size
TrafficGen	Full map	0.1602	0.1192	0.1189	0.0932
TraffusionGen	Full map + interaction	0.0634	0.0740	0.0697	0.0665
TrafficGen-WeakMap	Test-time map removal	0.3540	0.4361	0.4981	0.7428
TraffusionGen-WeakMap	Test-time map removal + interaction	0.6289	0.3322	0.3055	0.3650
MapLight-AgentInit	Map-light + interaction	0.1143 ± 0.0002	0.1144 ± 0.0004	0.1141 ± 0.0007	0.0734 ± 0.0002
InterInit	Strict no-map + interaction	0.2461 ± 0.0279	0.1147 ± 0.0001	0.1149 ± 0.0000	0.0797 ± 0.0078

The results provide a vertical view of traffic scenario initialization under decreasing map-prior availability. Under full-map conditions, TraffusionGen outperforms TrafficGen across all four attributes, indicating that enhanced interaction modeling is beneficial when complete map priors are available.

The two test-time explicit-map-removal baselines show that directly removing explicit HD-map inputs from map-conditioned models causes substantial degradation. TrafficGen-WeakMap degrades across all four attributes, with Speed, Position, Heading, and Size MMD increasing to 0.3540, 0.4361, 0.4981, and 0.7428, respectively. TraffusionGen-WeakMap also suffers from clear degradation after explicit map removal. Although its Position, Heading, and Size MMD values are lower than those of TrafficGen-WeakMap, its Speed MMD increases substantially to 0.6289. This suggests that interaction modeling can partially

stabilize some spatial attributes under test-time map removal, but it does not prevent severe speed-distribution mismatch when the model has been trained under full-map conditions.

Compared with the test-time explicit-map-removal baselines, MapLight-AgentInit achieves much lower MMD values. This is because MapLight-AgentInit is trained and evaluated under an explicit global map-removal protocol, rather than only removing the map input at inference time. However, MapLight-AgentInit still retains bounded local segment cues and should therefore be interpreted as a map-light setting rather than a strict no-map setting.

When the setting is further tightened from map-light initialization to strict no-map initialization, the degradation becomes attribute-asymmetric. Compared with MapLight-AgentInit, InterInit maintains similar Position and Heading MMD values, while its Speed MMD increases from 0.1143 to 0.2461. This indicates that strict removal of both explicit HD-map inputs and agent-internal road-related cues does not cause uniform degradation across all attributes. Instead, speed becomes the dominant failure dimension, whereas position and heading remain relatively stable. This observation motivates the following diagnostic analysis of the velocity output.

5.2. Speed-Magnitude Collapse Diagnosis

The global MMD comparison in Table 3 identifies speed as the dominant failure dimension under strict no-map removal. To further locate the source of this degradation, velocity is decomposed into speed magnitude and velocity direction. The diagnostic analysis shows that the mismatch is mainly caused by speed magnitude rather than velocity direction.

Figure 3 compares the mean and standard deviation of speed magnitude between the ground truth and InterInit. InterInit has a mean speed magnitude of 4.2337, which is close to the ground-truth mean of 4.2864. However, its speed-magnitude standard deviation is 0.0000, whereas the ground-truth standard deviation is 5.7902. This confirms that the major speed failure is not mean bias, but speed-magnitude variance collapse. In other words, under strict no-map removal, InterInit tends to predict an almost constant speed magnitude.

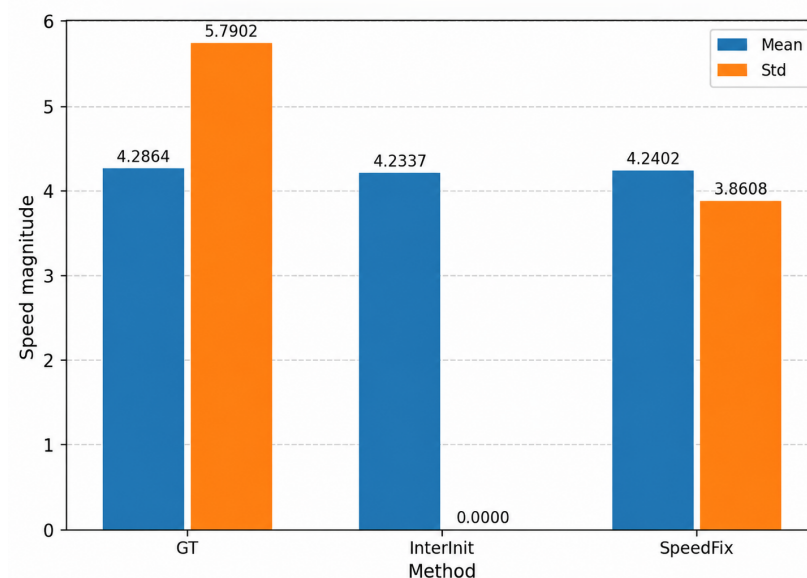


Figure 3. Speed-magnitude statistics for the ground truth, InterInit, and InterInit-SpeedFix. InterInit has a mean speed magnitude close to the ground truth but collapses to nearly zero variance. InterInit-SpeedFix restores a substantial portion of the missing speed-magnitude variability after post-diagnostic repair.

This phenomenon is consistent with the nature of speed generation. Speed magnitude is closely related to road geometry, lane type, traffic state, intersection structure, curvature, and local speed-limit priors. When both explicit maps and agent-internal road cues are removed, the model lacks these key constraints and tends to predict a conservative average-speed template.

5.3. Post-Diagnostic SpeedFix Repair

Based on the above diagnosis, InterInit-SpeedFix is introduced as a localized repair strategy for speed-magnitude collapse. Instead of redesigning the whole initialization framework, SpeedFix replaces deterministic speed regression with probabilistic modeling in log-speed space.

Let v denote speed magnitude. The log-speed variable is defined as

$$z = \log(v + \epsilon), \quad (16)$$

where ϵ is a small positive constant for numerical stability. The model predicts the mean and variance of z , and optimizes the negative log-likelihood objective:

$$\mathcal{L}_{\text{speed}}^{\log} = -\log \mathcal{N}(z; \mu_z, \sigma_z^2). \quad (17)$$

All other prediction heads and inference procedures remain unchanged. Therefore, SpeedFix directly targets the diagnosed failure mode rather than introducing a new full initialization architecture.

Table 4 reports the repaired results after applying SpeedFix. InterInit-SpeedFix reduces Speed MMD from 0.2461 to 0.1210, corresponding to a 50.9% improvement. Meanwhile, Position and Heading remain stable, and Size also remains comparable. This result verifies that the dominant speed failure can be repaired through localized probabilistic modeling after the failure mode has been identified.

Table 4. Effect of post-diagnostic SpeedFix repair. Lower MMD values are better.

Model	Setting	Speed	Position	Heading	Size
InterInit	Strict no-map	0.2461 ± 0.0279	0.1147 ± 0.0001	0.1149 ± 0.0000	0.0797 ± 0.0078
InterInit-SpeedFix	Strict no-map + SpeedFix	0.1210 ± 0.0005	0.1148 ± 0.0003	0.1146 ± 0.0002	0.0766 ± 0.0063

Figure 3 further shows that the predicted speed-magnitude standard deviation increases from 0.0000 to 3.8608 after applying SpeedFix. Although this value is still lower than the ground-truth standard deviation of 5.7902, it substantially mitigates the collapse observed in InterInit. Therefore, SpeedFix improves the speed distribution not by merely shifting the mean prediction, but by restoring part of the missing speed-magnitude variability.

To further examine the robustness of the SpeedFix improvement, we conducted seed-level and evaluation-record-level significance analyses, as summarized in Table 5. Across three random seeds, InterInit-SpeedFix significantly reduces Speed MMD compared with InterInit. A paired bootstrap analysis over 65,588 evaluation records from one representative seed further confirms that the reduction remains stable under evaluation-sample resampling. Following the TrafficGen evaluation implementation, Speed MMD is computed using the two-dimensional velocity vector $[v_x, v_y]$.

The seed-level paired test should be interpreted as evidence across independently trained runs, while the bootstrap analysis evaluates whether the observed reduction is stable with respect to the finite evaluation set. Although the number of training seeds is limited, both analyses lead to the same conclusion: the Speed MMD reduction remains positive and statistically reliable. In particular, the bootstrap confidence interval does

not overlap zero, indicating that the observed improvement is not caused by evaluation-sample fluctuation.

Table 5. Statistical significance analysis of the SpeedFix improvement on Speed MMD.

Analysis Level	Data Basis	InterInit	InterInit-SpeedFix	Statistical Evidence
Seed-level test	Three random seeds	0.2461 ± 0.0279	0.1210 ± 0.0005	Reduction = 0.1251 (50.83%); paired <i>t</i> -test, <i>t</i> = 7.79, <i>df</i> = 2, <i>p</i> = 0.0161
Evaluation-record bootstrap	65,588 records from one representative seed	0.2658	0.1206	Reduction = 0.1452; 95% CI [0.1437, 0.1466]; paired bootstrap, <i>p</i> < 0.001

We also compare the computational cost of InterInit and InterInit-SpeedFix. Both models use the same configuration with *hidden_dim* = 1024 and batch size 32. Since SpeedFix only changes the speed head from a one-dimensional deterministic output to a two-dimensional probabilistic output, it introduces only 257 additional parameters. As shown in Table 6, the model-forward overhead is 0.006 s per batch, or 0.1875 ms per sample, indicating that SpeedFix improves Speed MMD with only a small computational overhead.

Table 6. Computational cost of InterInit and InterInit-SpeedFix.

Model	Extra Params	Forward Time (s/Batch)	Forward Time (ms/Sample)
InterInit	–	0.036	1.1250
InterInit-SpeedFix	+257	0.042	1.3125

Note: Forward time is obtained from the PyTorch Lightning profiler (PyTorch 1.10.0 with CUDA 11.3; PyTorch Lightning 1.5.10) using the *model_forward* mean duration with batch size 32. The reported time measures model forward computation and excludes data loading, loss computation, backpropagation, optimizer updates, MMD computation, and geometric post-processing.

5.4. Ablation Study of SpeedFix

To identify the contribution of each component in SpeedFix, Table 7 compares deterministic raw-speed regression, deterministic log-speed regression, probabilistic raw-speed modeling, and the full probabilistic log-speed SpeedFix. The original InterInit suffers from severe speed-magnitude variance collapse, with a speed standard deviation of 0.0000. Log-Det only slightly reduces Speed MMD and still produces very limited speed variability. Raw-Prob substantially restores speed variance and reduces Speed MMD, showing that probabilistic modeling is the key factor for mitigating variance collapse. The full InterInit-SpeedFix achieves the best overall performance, with the lowest Speed MMD, mean absolute speed error, and P95 absolute speed error. These results indicate that log-speed transformation and probabilistic modeling are complementary, while probabilistic modeling plays the dominant role.

5.5. Density-Stratified Speed MMD

Figure 4 shows Speed MMD before and after SpeedFix across different agent-density buckets. The four density buckets are 8–10, 11–15, 16–20, and 21+ agents. The results show that SpeedFix consistently reduces Speed MMD in all density regimes.

Specifically, Speed MMD decreases from 0.3690 to 0.1816 in the 8–10 bucket, from 0.2747 to 0.1316 in the 11–15 bucket, from 0.2272 to 0.1008 in the 16–20 bucket, and from 0.2063 to 0.0769 in the 21+ bucket. This shows that the effect of SpeedFix is not restricted to a specific scenario complexity level, but is observed in sparse, medium-density, and dense scenes.

Table 7. Ablation study of SpeedFix components on the full evaluation set.

Model	Speed MMD	Speed Mean	Speed std	Mean Abs. Error	P95 Abs. Error
Ground Truth	Ref.	4.2864	5.7902	Ref.	Ref.
InterInit	0.2658	4.2337	0.0000	4.7278	11.8214
Log-Det	0.2352	4.2396	0.0360	4.3768	10.9786
Raw-Prob	0.1436	4.2358	3.6825	3.8907	9.0856
InterInit-SpeedFix	0.1206	4.2402	3.8608	3.1784	8.7901

Note: Ground Truth is used as the reference distribution. Therefore, Speed MMD and absolute-error metrics are not reported for it. All ablation variants are evaluated on the full evaluation set using the same representative seed and the same evaluation protocol. Speed MMD is computed using the two-dimensional velocity vector, while speed mean, speed standard deviation, and absolute-error statistics are computed from speed magnitude.

In addition, Speed MMD generally decreases as scene density increases. A possible reason is that denser scenes provide richer interaction context, enabling the model to obtain stronger constraints from surrounding agents. The relatively larger improvement in high-density scenarios suggests that interaction context and probabilistic speed modeling are complementary.

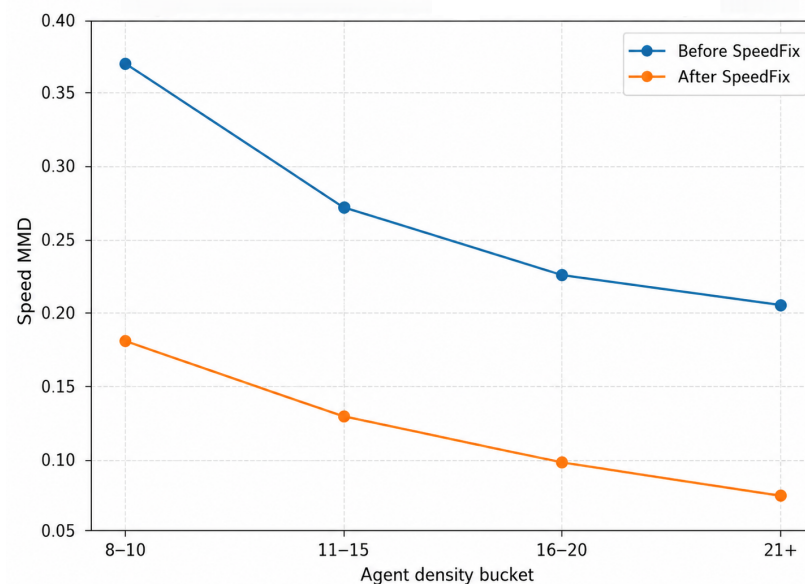


Figure 4. Speed MMD before and after SpeedFix across density buckets. SpeedFix consistently improves Speed MMD in all agent-density regimes.

5.6. Absolute Speed-Magnitude Error Statistics

Figure 5 reports absolute speed-magnitude error statistics before and after SpeedFix. The mean error decreases from 4.7278 to 3.1784, and the median error decreases from 4.2337 to 2.3724. The 90th percentile error decreases from 8.9317 to 7.0426, and the 95th percentile error decreases from 11.8214 to 8.7901.

These results show that SpeedFix improves not only average error but also typical and high-error cases. Together with the variance recovery shown in Figure 3, this indicates that SpeedFix does not simply shift the mean speed prediction. Instead, it improves the overall speed-magnitude distribution by restoring more reasonable variability and uncertainty.

5.7. Downstream Trajectory-Generation Validation

To examine whether the improved speed initialization propagates to subsequent trajectory generation, we compare InterInit and InterInit-SpeedFix using the same downstream trajectory-generation model, test split, rollout configuration, sampling protocol, and SCR

calculation procedure. The only changed factor is the initialization source. Mean ADE, Mean FDE, and Scenario Collision Rate (SCR) are reported.

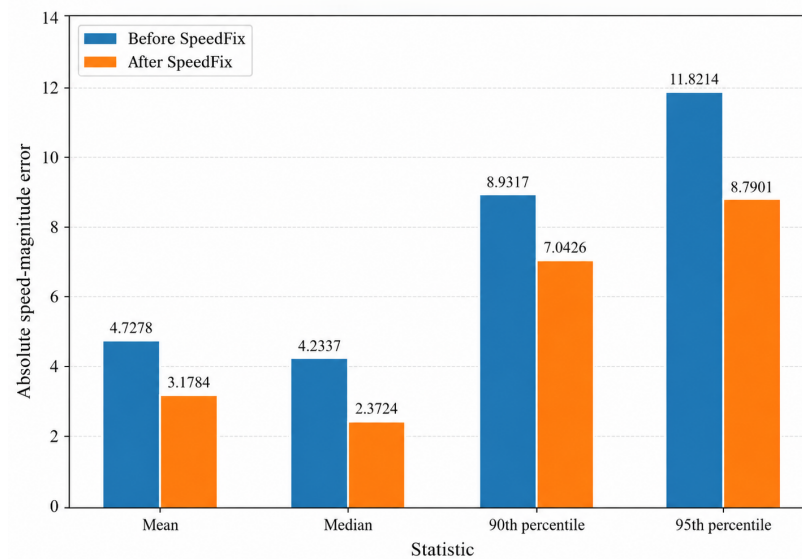


Figure 5. Absolute speed-magnitude error statistics before and after SpeedFix. SpeedFix reduces mean, median, 90th percentile, and 95th percentile errors, indicating improvement across both typical and high-error cases.

As shown in Table 8, InterInit-SpeedFix reduces Mean ADE from 2.52 m to 1.79 m and Mean FDE from 6.32 m to 5.07 m, corresponding to 29.0% and 19.8% relative reductions, respectively. SCR also decreases from 7.28% to 6.96%. These results indicate that SpeedFix not only improves the initial speed distribution but also benefits downstream trajectory-generation quality.

Table 8. Downstream trajectory-generation validation using InterInit and InterInit-SpeedFix initialization. Lower values are better.

Initialization Source	Mean ADE (m)	Mean FDE (m)	SCR (%)
InterInit	2.52	6.32	7.28
InterInit-SpeedFix	1.79	5.07	6.96
Reduction	29.0%	19.8%	0.32 pp

6. Discussion

6.1. Interpretation of Attribute-Asymmetric Degradation

The experimental results show that speed is more sensitive to strict map-prior removal than position and heading. A possible explanation is that spatial attributes can still be partially constrained by observed agent configurations, local density patterns, relative positions, and the fixed decoding-slot space. In contrast, speed magnitude depends more strongly on road semantics and local traffic rules, including road type, lane curvature, intersection structure, speed limits, lane topology, and traffic state. When both external HD maps and agent-internal road cues are removed, these constraints become unavailable, and the model tends to predict an average-speed template. This explains why the dominant failure mode appears as speed-magnitude variance collapse rather than as a large mean-speed bias.

This finding also clarifies the practical boundary of interaction-only initialization. Multi-agent interaction information can provide useful local relational cues, but it does not fully replace road-structure priors. Therefore, the proposed strict no-map setting

should be interpreted as a diagnostic framework for identifying which attributes remain recoverable from interaction-only information and which attributes require additional priors or targeted correction.

6.2. Baseline Positioning and Practical Implications

The baseline comparison is organized by information regimes rather than by forcing all models into an identical strict no-map interface. TrafficGen and TraffusionGen characterize full-map-prior performance without and with enhanced interaction modeling, respectively. The WeakMap variants measure test-time explicit-map-removal sensitivity of full-map-trained models, MapLight-AgentInit represents a train/test-consistent explicit-map-removed setting with local cues, and InterInit characterizes strict map-prior removal. This organization avoids conflating strict no-map initialization with artificially damaged versions of map-conditioned models.

The proposed framework is not intended to replace map-conditioned initialization when reliable HD maps are available. Its value lies in providing a controlled diagnosis-and-repair workflow for challenging map-prior conditions, such as missing, delayed, incomplete, or unreliable map information. The SpeedFix result further shows that a diagnosed failure mode does not always require a full architectural redesign. For speed-magnitude collapse, a localized probabilistic log-speed head substantially improves the speed distribution while preserving the original inference pipeline.

6.3. Limitations and Future Work

This study has several limitations. First, although the revised manuscript includes downstream trajectory-generation validation using ADE, FDE, and SCR, full closed-loop planning evaluation, map-based off-road analysis, and long-horizon interaction stability remain to be further explored. Second, SpeedFix mitigates speed-magnitude collapse but does not recover full road semantics. It should therefore be interpreted as a localized repair for a diagnosed speed failure mode, rather than as a replacement for reliable map priors.

Third, the experiments are conducted on the Waymo Open Motion Dataset. The proposed strict map-prior removal protocol is not tied to Waymo-specific HD-map parameters, because external map polylines, lane topology, and agent-internal road-related cues are removed from the model input. However, trained model weights are not assumed to transfer directly without adaptation, since different datasets and regions may have different coordinate conventions, annotation formats, agent taxonomies, road topologies, traffic rules, agent-density distributions, and driving styles. Cross-dataset validation on datasets such as nuScenes and Argoverse will require adapting the feature interface, applying the same road-related masking principle, and re-evaluating whether the identified speed-magnitude failure remains dominant under different regional driving styles. Future work will therefore conduct systematic cross-dataset validation and further investigate practical settings with online perception uncertainty, partial map fragments, and dynamically changing local road structures.

7. Conclusions

This study investigated traffic scenario initialization under strict map-prior removal. By masking both external HD-map inputs and agent-internal road-related cues, we constructed a controlled setting for analyzing the capability boundary of interaction-only initialization. The evaluation was organized as a progressive prior-removal analysis across full-map-prior, test-time explicit-map-removal, map-light, and strict no-map regimes.

The results show attribute-asymmetric degradation under strict map-prior removal. Position and heading remain relatively stable, whereas speed becomes the dominant failure

dimension. Diagnostic analysis further shows that the speed degradation is mainly caused by speed-magnitude collapse: InterInit predicts a mean speed close to the ground truth but with nearly zero variance.

Based on this diagnosis, InterInit-SpeedFix replaces deterministic speed regression with probabilistic log-speed modeling. Without changing the overall inference pipeline, it reduces Speed MMD from 0.2461 to 0.1210, corresponding to a 50.9% improvement, while maintaining stable position, heading, and size performance. Additional density-stratified, error-statistic, significance, efficiency, and downstream trajectory-generation analyses support the effectiveness and lightweight nature of the repair.

Overall, the study shows that interaction-only initialization can partially support spatial attributes under missing map priors, but speed modeling remains highly sensitive to road-related information and benefits from targeted uncertainty-aware correction.

Author Contributions: Conceptualization, K.Y. and S.Q.; methodology, K.Y.; software, K.Y.; validation, K.Y.; formal analysis, K.Y.; investigation, K.Y.; data curation, K.Y.; writing—original draft preparation, K.Y.; writing—review and editing, K.Y. and S.Q.; visualization, K.Y.; supervision, S.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study were obtained from the Waymo Open Motion Dataset. Restrictions apply to the availability and redistribution of these data, which were used under the terms of the Waymo Open Dataset license. The dataset is publicly available from the Waymo Open Dataset platform, and users should obtain it directly from Waymo Open Dataset.

Acknowledgments: The authors would like to thank the providers of the Waymo Open Motion Dataset for making the dataset publicly available for research purposes.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HD	High-definition
MMD	Maximum Mean Discrepancy
WOMD	Waymo Open Motion Dataset
NLL	Negative log-likelihood

References

1. Tan, S.; Wong, K.; Wang, S.; Manivasagam, S.; Ren, M.; Urtasun, R. SceneGen: Learning to Generate Realistic Traffic Scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 892–901.
2. Feng, L.; Li, Q.; Peng, Z.; Tan, S.; Zhou, B. TrafficGen: Learning to Generate Diverse and Realistic Traffic Scenarios. In Proceedings of the IEEE International Conference on Robotics and Automation, London, UK, 29 May–2 June 2023.
3. Mahjourian, R.; Mu, R.; Likhoshesterov, V.; Mouglin, P.; Huang, X.; Messias, J.; Whiteson, S. UniGen: Unified Modeling of Initial Agent States and Trajectories for Generating Autonomous Driving Scenarios. In Proceedings of the IEEE International Conference on Robotics and Automation, Yokohama, Japan, 13–17 May 2024.
4. Tan, S.; Ivanovic, B.; Weng, X.; Pavone, M.; Krahenbuhl, P. Language Conditioned Traffic Generation. In *Proceedings of the Conference on Robot Learning, Atlanta, GA, USA, 6–9 November 2023*; PMLR: Cambridge, MA, USA, 2023; pp. 2714–2752.
5. Lu, J.; Wong, K.; Zhang, C.; Suo, S.; Urtasun, R. SceneControl: Diffusion for Controllable Traffic Scene Generation. In Proceedings of the IEEE International Conference on Robotics and Automation, Yokohama, Japan, 13–17 May 2024; pp. 16908–16914.
6. Asrat, K.T.; Cho, H.J. A Comprehensive Survey on High-Definition Map Generation and Maintenance. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 232. [[CrossRef](#)]

7. Elghazaly, G.; Frank, R.; Harvey, S.; Safko, S. High-Definition Maps: Comprehensive Survey, Challenges and Future Perspectives. *IEEE Open J. Intell. Transp. Syst.* **2023**, *4*, 527–550. [[CrossRef](#)]
8. Zang, A.; Xu, R.; Zhou, F. Data Issues in High-Definition Maps Furniture: A Survey. *ACM Trans. Spat. Algorithms Syst.* **2024**, *10*, 1–37. [[CrossRef](#)]
9. Hui, N.; Jiang, Z.; Cai, Z.; Ying, S. Vision-HD: Road Change Detection and Registration Using Images and High-Definition Maps. *Int. J. Geogr. Inf. Sci.* **2024**, *38*, 454–477. [[CrossRef](#)]
10. Rowe, L.; Girgis, R.; Gosselin, A.; Paull, L.; Pal, C.; Heide, F. Scenario Dreamer: Vectorized Latent Diffusion for Generating Driving Simulation Environments. *arXiv* **2025**, arXiv:2503.22496.
11. Chang, Y.; Xi, C.; Luo, Z. Specific Scenario Generation Method for Trustworthiness Testing of Autonomous Vehicles Based on Interaction Coding. *Appl. Sci.* **2025**, *15*, 10656. [[CrossRef](#)]
12. Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; Alahi, A. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 18–22 June 2018.
13. Salzmann, T.; Ivanovic, B.; Chakravarty, P.; Pavone, M. Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data. In *Computer Vision—ECCV 2020*; Springer: Berlin/Heidelberg, Germany, 2020.
14. Yuan, Y.; Weng, X.; Ou, Y.; Kitani, K. AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021.
15. Gao, J.; Ma, C.; Li, M.; Dai, Y.; Zhu, Y.; Wang, Y.; Yang, B. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020.
16. Liang, M.; Yang, B.; Hu, R.; Wang, Y.; Urtasun, R. Learning Lane Graph Representations for Motion Forecasting. In *Computer Vision—ECCV 2020*; Springer: Berlin/Heidelberg, Germany, 2020.
17. Ettinger, S.; Cheng, S.; Caine, B.; Liu, C.; Zhao, H.; Pradhan, S.; Chai, Y.; Sapp, B.; Qi, C.R.; Zhou, Y.; et al. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.